



Large Deviations for a matching problem related to the ∞ -Wasserstein distance

José Trashorras

► To cite this version:

José Trashorras. Large Deviations for a matching problem related to the ∞ -Wasserstein distance. 2011. hal-00641378v1

HAL Id: hal-00641378

<https://hal.science/hal-00641378v1>

Preprint submitted on 15 Nov 2011 (v1), last revised 15 Feb 2017 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Large Deviations for a matching problem related to the ∞ -Wasserstein distance

José Trashorras

*Université Paris-Dauphine
Ceremade
Place du maréchal de Lattre de Tassigny
75775 Paris Cedex 16 France
e-mail: xose@ceremade.dauphine.fr*

Abstract: Let (E, d) be a compact metric space, $X = (X_1, \dots, X_n, \dots)$ and $Y = (Y_1, \dots, Y_n, \dots)$ two independent sequences of independent E -valued random variables and $(L_n^X)_{n \geq 1}$ and $(L_n^Y)_{n \geq 1}$ the associated sequences of empirical measures. We establish a Large Deviations Principle for $(W_\infty(L_n^X, L_n^Y))_{n \geq 1}$ where W_∞ is the ∞ -Wasserstein distance.

AMS 2000 subject classifications: Primary 60F10, 60C05; secondary 60D05, 05B99..

Keywords and phrases: Large Deviations, ∞ -Wasserstein distance, Minimax matching problem..

1. Introduction and results

We say that a sequence of Borel probability measures $(P^n)_{n \geq 1}$ on a topological space \mathcal{Y} obeys a Large Deviation Principle (hereafter abbreviated LDP) with rate function I if I is a non-negative, lower semi-continuous function defined on \mathcal{Y} such that

$$-\inf_{y \in A^\circ} I(y) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log P^n(A) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log P^n(A) \leq -\inf_{y \in \bar{A}} I(y)$$

for any measurable set $A \subset \mathcal{Y}$, whose interior is denoted by A° and closure by \bar{A} . If the level sets $\{y : I(y) \leq \alpha\}$ are compact for every $\alpha < \infty$, I is called a good rate function. With a slight abuse of language we say that a sequence of random variables obeys an LDP when the sequence of measures induced by these random variables obeys an LDP. For a background on the theory of large deviations see Dembo and Zeitouni [4] and references therein.

Let (E, d) be a metric space. In recent years there has been a lot of interest in considering the space $M^1(E)$ of Borel probability measures on E endowed

with the so-called p -Wasserstein distances

$$W_p(\nu, \gamma) = \inf_{Q \in \mathcal{C}(\nu, \gamma)} \left\{ \left(\int_{E \times E} d(x, y)^p Q(dx, dy) \right)^{1/p} \right\}$$

where $p \in [1, \infty)$ and $\mathcal{C}(\nu, \gamma)$ stands for the set of Borel probability measures on E^2 with first marginal $Q_1 = \nu$ and second marginal $Q_2 = \gamma$, see Chapter 6 in [13] for a broad review. However, the ∞ -Wasserstein distance

$$W_\infty(\nu, \gamma) = \inf_{Q \in \mathcal{C}(\nu, \gamma)} \sup \mathcal{S}(Q \circ d^{-1})$$

where $\mathcal{S}(Q \circ d^{-1})$ stands for the support of the probability measure $Q \circ d^{-1}$ has attracted much less attention despite the fact that, in words of [13], "this distance is useful in a surprising number of problems" (page 109).

Our framework is the following: We are given a compact metric space (E, d) . Without loss of generality we can assume that $\sup_{x, y \in E} d(x, y) = 1$. Let $(X_n)_{n \geq 1}$ and $(Y_n)_{n \geq 1}$ be two independent sequences of E -valued independent random variables defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We assume that all the X_i 's (resp. Y_i 's) have the same distribution μ^1 (resp. μ^2). For every $n \geq 1$ we consider the empirical measures

$$L_n^X = \frac{1}{n} \sum_{i=1}^n \delta_{X_i} \quad \text{and} \quad L_n^Y = \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}.$$

In [7] Ganesh and O'Connell conjecture that $(W_\infty(L_n^X, L_n^Y))_{n \geq 1}$ obeys an LDP with rate function

$$I_\infty(x) = \inf_{\substack{\nu^1, \nu^2 \in M^1(E) \\ W_\infty(\nu^1, \nu^2) = x}} \{H(\nu^1 | \mu^1) + H(\nu^2 | \mu^2)\}$$

where for any two $\nu, \mu \in M^1(E)$

$$H(\nu | \mu) = \begin{cases} \int_E \log \frac{d\nu}{d\mu} d\nu & \text{if } \nu \ll \mu \\ \infty & \text{otherwise.} \end{cases}$$

If instead of I_∞ we consider its lower semi-continuous regularization J_∞ (see e.g. Chapter 1 in [10]) which is defined by

$$J_\infty(x) = \sup_{\delta > 0} \inf_{y \in B(x, \delta)} I_\infty(y)$$

where $B(x, \delta)$ stands for the open ball centered at x with radius δ , then we can prove that

Theorem 1.1. *The sequence $(W_\infty(L_n^X, L_n^Y))_{n \geq 1}$ satisfies an LDP on $[0, 1]$ with good rate function $J_\infty(x)$.*

Due to the highly discontinuous nature of W_∞ we can not determine if I_∞ and J_∞ are equal or not. Nevertheless, the (apparently) more involved nature of the definition of J_∞ is not an obstacle when deriving its properties, see below.

Since for every $n \geq 1$ every $Q \in \mathcal{C}(L_n^X, L_n^Y)$ can be represented as a bi-stochastic matrix and since, according to the Birkoff-Von Neumann Theorem, every bi-stochastic matrix is a convex combination of permutation matrices (see e.g. Theorem 5.5.1 in [11]) we have

$$W_\infty(L_n^X, L_n^Y) = \min_{\sigma \in \mathfrak{S}_n} \max_{1 \leq i \leq n} d(X_i, Y_{\sigma(i)}) \quad (1.1)$$

where \mathfrak{S}_n stands for the set of permutations of $\{1, \dots, n\}$. Hence, computing $W_\infty(L_n^X, L_n^Y)$ is nothing but solving a minimax matching problem which is a fundamental combinatorial question with many potential applications in seemingly unrelated areas. As it is well illustrated in [8] the statistic $W_\infty(L_n^X, L_n^Y)$ is connected with e.g. maximum up-right matching problems, dynamic allocation, wafer-scale integration of systolic arrays, testing pseudo-random number generators etc...

So far, results on $(W_\infty(L_n^X, L_n^Y))_{n \geq 1}$ have been focused on independent $[0, 1]^2$ -valued X_i 's and Y_i 's with common distribution $\mu^1 = \mu^2 = \lambda$ where λ stands for the uniform distribution over $[0, 1]^2$. In [8] Leighton and Shor establish that there exists a $K > 0$ such that

$$\frac{1}{K} n^{-1/2} (\log n)^{3/4} \leq W_\infty(L_n^X, L_n^Y) \leq K n^{-1/2} (\log n)^{3/4}$$

with probability $1 - o(1)$. Using majorizing measures, Talagrand showed in [12] that the latter result is a particular case of a general property of empirical discrepancies. In [7] the authors prove that $(W_\infty(L_n^X, \lambda))_{n \geq 1}$ obeys an LDP with good rate function

$$J_\infty(x) = \inf_{\substack{\nu \in M^1(E) \\ W_\infty(\nu, \lambda) = x}} \{H(\nu|\lambda)\}.$$

To obtain this result they prove that in this particular setting the map $\nu \mapsto W_\infty(\nu, \lambda)$ is continuous with respect to the weak convergence topology, and then apply the contraction principle, see Theorem 4.2.1 in [4]. They further show in the same framework as here, i.e. E -valued X_i 's and Y_i 's that $(W_1(L_n^X, L_n^Y))_{n \geq 1}$ obeys an LDP with good rate function

$$I_1(x) = \inf_{\substack{\nu^1, \nu^2 \in M^1(E) \\ W_1(\nu^1, \nu^2) = x}} \{H(\nu^1|\mu^1) + H(\nu^2|\mu^2)\}.$$

Their proof relies on the fact that, according to Kantorovitch-Rubinstein Theorem (see e.g. Theorem 11.8.2 in [5]), when E is compact W_1 generates the weak convergence topology. As a consequence $(L_n^X)_{n \geq 1}$ and $(L_n^Y)_{n \geq 1}$ satisfy an LDP

on $M^1(E)$ endowed with W_1 and again the contraction principle leads to an LDP for $(W_1(L_n^X, L_n^Y))_{n \geq 1}$. Following the same approach, one can deduce from Theorem 1.1 in [14] that for every $p \in [1, \infty)$ the sequence $(W_p(L_n^X, L_n^Y))_{n \geq 1}$ obeys an LDP with good rate function

$$I_p(x) = \inf_{\substack{\nu^1, \nu^2 \in M^1(E) \\ W_p(\nu^1, \nu^2) = x}} \{H(\nu^1 | \mu^1) + H(\nu^2 | \mu^2)\}.$$

One might wonder why this idea is not applicable to the analysis of the LD properties of $(W_\infty(L_n^X, L_n^Y))_{n \geq 1}$. Actually an LDP for $(L_n^X)_{n \geq 1}$ can not hold when $M^1(E)$ is endowed with W_∞ at least at this level of generality. Indeed, consider the probability measure μ^1 on $E = [0, 4]$ which density with respect to the Lebesgue measure is $f(x) = 1/2(1_{[0,1]}(x) + 1_{[3,4]}(x))$ and assume that $(L_n^X)_{n \geq 1}$ satisfies an LDP on $(M^1(E), W_\infty)$ with some rate function \mathcal{R} . Clearly for every odd integer n we have $\mathbb{P}(W_\infty(L_n^X, \mu^1) < 3/2) = 0$ hence for every $\nu \in B(\mu, 3/2)$ we necessarily have $\mathcal{R}(\nu) = \infty$. So we should get

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, \mu^1) \leq 1) = -\infty \quad (1.2)$$

but for every even n we have $\mathbb{P}(W_\infty(L_n^X, \mu^1) \leq 1) \geq C_n^{n/2} 2^{-n}$ which contradicts (1.2).

Finally lets us give some facts about J_∞ . For simplicity we assume that $\mu^1 = \mu^2 = \mu$. For any two different $a_i, a_j \in \mathcal{S}(\mu)$ we write $a_i \leftrightarrow a_j$ if and only if for every integer $L \geq 3$ and every $(\alpha_1, \dots, \alpha_L) \in \mathcal{S}(\mu)^L$ such that $\alpha_1 = a_i$ and $\alpha_L = a_j$ there exists at least one $\alpha_s, 1 \leq s \leq L-1$ such that $d(\alpha_s, \alpha_{s+1}) \geq d(a_i, a_j)$. In words $a_i \leftrightarrow a_j$ means that one can not decompose a mass transport from a_i to a_j into stages starting/ending on elements of $\mathcal{S}(\mu)$ that are all of length strictly smaller than $d(a_i, a_j)$. In particular for every $a_i \in \mathcal{S}(\mu)$ we have $a_i \leftrightarrow a_i$ and we shall write $a_i \nleftrightarrow a_j$ when the above property is not satisfied. Consider

$$\mathcal{Z}_\mu = \{x \in [0, 1] : \text{there exists } a_i, a_j \in \mathcal{S}(\mu) \text{ such that } a_i \leftrightarrow a_j \text{ and } x = d(a_i, a_j)\}.$$

The highly discontinuous nature of W_∞ is responsible for the unusual behavior of J_∞ when μ has finite support.

Proposition 1.1. *If $\mathcal{S}(\mu) = \{a_1, \dots, a_N\}$ then $J_\infty(x) = 0$ if and only if $x \in \mathcal{Z}_\mu$.*

It is noticeable (although not surprising in view of [2]) that the set of zeroes of the rate function depends on the support of μ but not on its exact density. We further show that

Proposition 1.2. *If $\mathcal{S}(\mu)$ is a connected subset of \mathbb{R}^d then $J_\infty(x) = 0$ if and only if $x = 0$.*

The paper is structured as follows : Section 2 is devoted to the proof of Theorem 1.1 while Section 3 deals with the proof of some complementary results.

2. Proof of Theorem 1.1

The basic idea in the proof of Theorem 1.1 is to partition E in such a way that we are essentially led to consider finite set valued X_i 's and Y_i 's. Indeed, on the one hand we will see that W_∞ is well-behaved with respect to partitioning (see Lemma 2.3 below) and on the other hand proceeding this way reduces the computation of probabilities to simple classical combinatorial estimates. In order to go from the particular case, i.e. E finite, to the general one we shall need some results on the weak convergence of nets of probability measures. In Section 2.1, after generalizing some properties of W_∞ from [2], we give an account on partitions, nets and the weak convergence topology. The LD lower bound is established in Section 2.2 while the LD upper bound is derived in Section 2.3.

2.1. Some preliminary facts

2.1.1. About W_∞

Here we collect some definitions and results that generalize results from [2] to compact metric spaces. Indeed, the latter reference is only concerned with measures supported on \mathbb{R}^d . These generalizations do not require any new idea, which is the reason why we postpone the corresponding proofs until Section 3.

Definition 2.1. *A probability measure $P \in M^1(E^2)$ is called infinitely cyclically monotone if and only if for every integer $n \geq 2$, every $(x_1, y_1), \dots, (x_n, y_n) \in \mathcal{S}(P)$ and every $\sigma \in \mathfrak{S}_n$ we have*

$$\max_{1 \leq i \leq n} d(x_i, y_i) \leq \max_{1 \leq i \leq n} d(x_i, y_{\sigma(i)}).$$

Infinitely cyclically monotone probability measures are the right couplings to consider when computing W_∞ distances as the following results show

Lemma 2.1. *(From Proposition 2.1 and Theorem 3.2 in [2]) For any two $\gamma, \nu \in M^1(E)$ there exists an infinitely cyclically monotone $P \in \mathcal{C}(\gamma, \nu)$ such that*

$$W_\infty(\gamma, \nu) = \sup \mathcal{S}(P \circ d^{-1}).$$

Lemma 2.2. *(From Theorem 3.4 in [2]) Any infinitely cyclically monotone $P \in M^1(E^2)$ satisfies*

$$W_\infty(P_1, P_2) = \sup \mathcal{S}(P \circ d^{-1}).$$

2.1.2. About partitions of E , nets and the weak convergence topology

Let \mathcal{P} be the set of finite measurable partitions of E into non-empty sets. To every $\Pi = (A_1, \dots, A_L) \in \mathcal{P}$ we associate once for all through the paper a family

$(s_1, \dots, s_L) \in E^L$ such that for every $1 \leq i \leq L$ we have $s_i \in A_i$. We further associate to every $\Pi = (A_1, \dots, A_L) \in \mathcal{P}$ a map π as follows

$$\begin{aligned} \pi: M^1(E) &\rightarrow M^1(\{s_1, \dots, s_L\}) \\ \nu &\mapsto \sum_{i=1}^L \nu(A_i) \delta_{s_i}. \end{aligned}$$

Finally for every $\Pi = (A_1, \dots, A_L) \in \mathcal{P}$ we define its maximal diameter as

$$\Delta(\Pi) = \max_{1 \leq i \leq L} \sup_{x, y \in A_i} d(x, y).$$

The following result links the W_∞ distance between elements of $M^1(E)$ to the analogue W_∞ distance between their contractions through π .

Lemma 2.3. *For every $\nu^1, \nu^2 \in M^1(E)$ and every $\Pi \in \mathcal{P}$ we have*

$$|W_\infty(\nu^1, \nu^2) - W_\infty(\pi(\nu^1), \pi(\nu^2))| \leq 2\Delta(\Pi).$$

As for the results of Section 2.1.1 the proofs of the lemmas in the present section are postponed to Section 3.

Next we give some more definitions and notations. Let $\Pi = (A_1, \dots, A_L)$ and $\mathcal{K} = (B_1, \dots, B_R)$ be two elements of \mathcal{P} . We say that \mathcal{K} is a *refinement* of Π if and only if for every $1 \leq i \leq L$, there exists $J_i \subset \{1, \dots, R\}$ such that $A_i = \cup_{j \in J_i} B_j$, and we denote it by $\Pi \preceq \mathcal{K}$. This makes (\mathcal{P}, \preceq) a directed set. Let us recall that (J, \preceq) is a *directed set* if and only if J is a non-empty set and \preceq is a reflexive and transitive relation on J such that for every $i, j \in J$ there exists a $k \in J$ such that $i \preceq k$ and $j \preceq k$. We introduce a general directed set (J, \preceq) since we will use both (\mathcal{P}, \preceq) and (\mathbb{N}, \leq) as directed sets and we do not want to be too specific in the results below. We call *net* any map $(P^j)_{j \in J}$ defined on a directed set (J, \preceq) . A topological space (T, \mathcal{T}) -valued net $(P^j)_{j \in J}$ is said to *converge* to some $P \in T$ if and only if for every neighborhood U of P there exists a $j(U) \in J$ such that for every $k \in J$ satisfying $j(U) \preceq k$ we have $P^k \in U$. We shall denote this $\lim_{j \in J} P^j = P$. We call *subnet* of $(P^j)_{j \in J}$ any sub-family $(P^l)_{l \in L}$ parametrized by a *cofinal* L , i.e. a subset $L \subset J$ such that for every $j \in J$ there exists $l \in L$ satisfying $j \preceq l$. (While this is not the most general definition of a subnet it will be sufficient for the problem considered here). Let us recall that a topological space (T, \mathcal{T}) is compact if and only if every net in T admits a subnet that converges to some point in T . Finally, for every real-valued net $(P^j)_{j \in J}$ we shall consider as usual

$$\limsup_{j \in J} P^j = \inf_{j \in J} \sup_{i: j \preceq i} P^i \text{ and } \liminf_{j \in J} P^j = \sup_{j \in J} \inf_{i: j \preceq i} P^i.$$

For this and other questions related to nets we refer to e.g. Chapter 8 in [9]. The following lemma is a consequence of a kind of Portmanteau result for nets of probability measures that will be derived in Section 3.

Lemma 2.4. *Every net $(P^j)_{j \in J}$ of probability measures supported on \mathbb{R} that converges weakly to some probability measure P satisfies*

$$\limsup_{j \in J} \mathcal{S}(P^j) \geq \sup \mathcal{S}(P). \quad (2.1)$$

2.2. LD lower bound

For every integer $n \geq 1$ we consider

$$M^{1,n}(E) = \left\{ \nu \in M^1(E) : \text{there exists } (x_1, \dots, x_n) \in E^n \text{ such that } \nu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \right\}.$$

The following lemma is the key point in the proof of the LD lower bound

Lemma 2.5. *For every $\varepsilon > 0$, every $\nu^1, \nu^2 \in M^1(E)$ and every $\Pi = (A_1, \dots, A_L) \in \mathcal{P}$ with $\Delta(\Pi) \leq \varepsilon/2$ there exists two sequences $(\nu^{1,n})_{n \geq 1}$ and $(\nu^{2,n})_{n \geq 1}$ such that*

1. *For every $n \geq 1$ we have $\nu^{1,n}, \nu^{2,n} \in M^{1,n}(E)$.*
2. *For every $1 \leq i \leq L$ we have $\nu^{1,n}(A_i) \rightarrow \nu^1(A_i)$ and $\nu^{2,n}(A_i) \rightarrow \nu^2(A_i)$.*
3. *There exists an N_0 such that for every $n \geq N_0$ we have*

$$|W_\infty(\nu^1, \nu^2) - W_\infty(\nu^{1,n}, \nu^{2,n})| \leq \varepsilon.$$

Proving Lemma 2.5 actually requires one more lemma

Lemma 2.6. *For every $Q \in M^1(E^2)$ and every $\Pi = (A_1, \dots, A_L) \in \mathcal{P}$ there exists a sequence $(Q^n = \frac{1}{n} \sum_{i=1}^n \delta_{(x_i^n, y_i^n)})_{n \geq 1}$ of elements of $M^1(E^2)$ such that*

1. *For every $1 \leq u, v \leq L$ we have $Q^n(A_u \times A_v) \rightarrow Q(A_u \times A_v)$.*
2. *The sequence $(Q^n)_{n \geq 1}$ converges weakly to Q .*
3. *If $Q(A_u \times A_v) = 0$ then for every $n \geq 1$ we have $Q^n(A_u \times A_v) = 0$.*

Proof of Lemma 2.6 Let Q be an element of $M^1(E^2)$ and $\Pi = (A_1, \dots, A_L) \in \mathcal{P}$. Let $(X_i, Y_i)_{i \leq 1}$ be a sequence of independent and Q -identically distributed random couples. According to the Strong Law of Large Numbers there exists an event B of probability 1 such that for every $1 \leq u, v \leq L$

$$\frac{1}{n} \sum_{i=1}^n 1_{A_u \times A_v}(X_i, Y_i) \rightarrow Q(A_u \times A_v)$$

on B . Moreover, since E^2 is separable (it is compact), according to Varadarajan's Lemma, see e.g. Theorem 11.4.1 in [5], there exists an event C of probability 1 such that

$$\frac{1}{n} \sum_{i=1}^n \delta_{(X_i, Y_i)} \xrightarrow{w} Q$$

on C where \xrightarrow{w} stands for the weak convergence of probability measures. Hence almost every realization of $\frac{1}{n} \sum_{i=1}^n \delta_{(X_i, Y_i)}$ can play the role of Q^n and the conclusions of Lemma 2.6 follow. \square

Proof of Lemma 2.5 Let ν^1 and ν^2 be two elements of $M^1(E)$, $\varepsilon > 0$ and $\Pi \in \mathcal{P}$ with $\Delta(\Pi) \leq \varepsilon/2$. According to Lemma 2.1 there is a $Q \in \mathcal{C}(\nu^1, \nu^2)$ such that

$$W_\infty(\nu^1, \nu^2) = \sup \mathcal{S}(Q \circ d^{-1}).$$

Let $(Q^n)_{n \geq 1}$ be the sequence of elements of $M^1(E^2)$ associated to Q and Π by Lemma 2.6. We shall prove that $(\nu^{1,n} = Q_1^n)_{n \geq 1}$ and $(\nu^{2,n} = Q_2^n)_{n \geq 1}$ meet the conditions of Lemma 2.5. First, (1) and (2) in Lemma 2.5 are clearly satisfied with these $(\nu^{1,n})_{n \geq 1}$ and $(\nu^{2,n})_{n \geq 1}$. Now, due to the definition of W_∞ , for every $n \geq 1$ we have

$$W_\infty(\nu^{1,n}, \nu^{2,n}) \leq \sup \mathcal{S}(Q^n \circ d^{-1})$$

and due to (1) and (3) in Lemma 2.6 there exists an N_1 such that for every $n \geq N_1$

$$\sup \mathcal{S}(Q^n \circ d^{-1}) \leq \sup \mathcal{S}(Q \circ d^{-1}) + \varepsilon$$

hence $W_\infty(\nu^{1,n}, \nu^{2,n}) \leq W_\infty(\nu^1, \nu^2) + \varepsilon$. Thus we are left to prove that there exists an N_2 such that for every $n \geq N_2$ we have $W_\infty(\nu^1, \nu^2) - \varepsilon \leq W_\infty(\nu^{1,n}, \nu^{2,n})$. Let us assume that this is not true i.e. that there exists a sequence $(n_k)_{k \geq 1}$ such that for every $k \geq 1$ we have

$$W_\infty(\nu^1, \nu^2) - \varepsilon > W_\infty(\nu^{1,n_k}, \nu^{2,n_k}). \quad (2.2)$$

According to Lemma 2.1 for every $k \geq 1$ there exists a $C^{n_k} \in \mathcal{C}(\nu^{1,n_k}, \nu^{2,n_k})$ such that

$$W_\infty(\nu^{1,n_k}, \nu^{2,n_k}) = \sup \mathcal{S}(C^{n_k} \circ d^{-1}).$$

The sequence $(C^{n_k})_{k \geq 1}$ admits a weakly converging subsequence since $M^1(E^2)$ is compact for the weak convergence topology. By a slight abuse of notation we still denote $(C^{n_k})_{k \geq 1}$ this converging subsequence. Let C be its limit. Due to (2) in Lemma 2.6 we necessarily have $C_1 = \nu^1$ and $C_2 = \nu^2$. Moreover $(C^{n_k} \circ d^{-1})_{k \geq 1}$ is a sequence of probability measures on \mathbb{R} that weakly converges to $C \circ d^{-1}$. Combining (2.2) with Lemma 2.4 we have

$$\begin{aligned} W_\infty(\nu^1, \nu^2) - \varepsilon &\geq \limsup_{k \rightarrow \infty} W_\infty(\nu^{1,n_k}, \nu^{2,n_k}) \\ &= \limsup_{k \rightarrow \infty} \sup \mathcal{S}(C^{n_k} \circ d^{-1}) \\ &\geq \sup \mathcal{S}(C \circ d^{-1}) \\ &\geq W_\infty(\nu^1, \nu^2) \end{aligned}$$

which can not be. The announced result follows. \square

Proof of the LD lower bound. As usual, in order to prove the LD lower bound it is sufficient to prove that for every $x \in [0, 1]$ and every $\varepsilon > 0$ we have

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in]x - \varepsilon, x + \varepsilon]) \geq -J_\infty(x). \quad (2.3)$$

In particular we can assume that $J_\infty(x) < \infty$ for otherwise (2.3) trivially holds. If $J_\infty(x) < \infty$ then for every $m \geq 1$ there exists an y_m such that $|x - y_m| < 1/m$, $I_\infty(y_m) < \infty$ and $\lim_{m \rightarrow \infty} I_\infty(y_m) = J_\infty(x)$. Now let us assume

that for every m such that $I_\infty(y_m) < \infty$ and every $\nu^1, \nu^2 \in M^1(E)$ such that $W_\infty(\nu^1, \nu^2) = y_m$ we have

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in]y_m - \varepsilon/4, y_m + \varepsilon/4]) \geq -H(\nu^1 | \mu^1) - H(\nu^2 | \mu^2). \quad (2.4)$$

Then for every m large enough

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in]x - \varepsilon, x + \varepsilon]) &\geq \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in]y_m - \varepsilon/4, y_m + \varepsilon/4]) \\ &\geq -I_\infty(y_m) \end{aligned}$$

which leads to (2.3) by letting $m \rightarrow \infty$. Hence it is sufficient to establish (2.4) and we will just do that.

Let $\varepsilon > 0$, $y_m \in [0, 1]$ be such that $I_\infty(y_m) < \infty$ and $\nu^1, \nu^2 \in M^1(E)$ be such that $W_\infty(\nu^1, \nu^2) = y_m$. According to Lemma 2.5 to any $\Pi = (A_1, \dots, A_L) \in \mathcal{P}$ with $\Delta(\Pi) \leq \varepsilon/32$ we can associate two sequences $(\nu^{1,n})_{n \geq 1}$ and $(\nu^{2,n})_{n \geq 1}$ and an integer N_0 such that for every $n \geq 1$ we have $\nu^{1,n}, \nu^{2,n} \in M^{1,n}(E)$ and for every $n \geq N_0$ one has

$$\begin{aligned} \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in]y_m - \varepsilon/4, y_m + \varepsilon/4]) &= \mathbb{P}(|W_\infty(L_n^X, L_n^Y) - W_\infty(\nu^1, \nu^2)| < \varepsilon/4) \\ &\geq \mathbb{P}(|W_\infty(L_n^X, L_n^Y) - W_\infty(\nu^{1,n}, \nu^{2,n})| \leq \varepsilon/8). \end{aligned}$$

We obviously have

$$\begin{aligned} |W_\infty(L_n^X, L_n^Y) - W_\infty(\nu^{1,n}, \nu^{2,n})| &\leq |W_\infty(L_n^X, L_n^Y) - W_\infty(\pi(L_n^X), \pi(L_n^Y))| + \\ &\quad + |W_\infty(\pi(L_n^X), \pi(L_n^Y)) - W_\infty(\pi(\nu^{1,n}), \pi(\nu^{2,n}))| + \\ &\quad + |W_\infty(\pi(\nu^{1,n}), \pi(\nu^{2,n})) - W_\infty(\nu^{1,n}, \nu^{2,n})| \end{aligned}$$

so due to Lemma 2.3 we get

$$\{\pi(L_n^X) = \pi(\nu^{1,n})\} \cap \{\pi(L_n^Y) = \pi(\nu^{2,n})\} \subset \{|W_\infty(L_n^X, L_n^Y) - W_\infty(\nu^{1,n}, \nu^{2,n})| \leq \varepsilon/8\}$$

hence

$$\begin{aligned} \mathbb{P}(|W_\infty(L_n^X, L_n^Y) - W_\infty(\nu^{1,n}, \nu^{2,n})| \leq \varepsilon/8) &\geq \mathbb{P}(\{\pi(L_n^X) = \pi(\nu^{1,n})\} \cap \{\pi(L_n^Y) = \pi(\nu^{2,n})\}) \\ &= \mathbb{P}(\{\pi(L_n^X) = \pi(\nu^{1,n})\}) \mathbb{P}(\{\pi(L_n^Y) = \pi(\nu^{2,n})\}) \end{aligned}$$

since the sequences $(X_n)_{n \geq 1}$ and $(Y_n)_{n \geq 1}$ are independent. It follows from elementary combinatorics (see e.g. Lemma 2.1.9 in [4]) that

$$\mathbb{P}(\{\pi(L_n^X) = \pi(\nu^{1,n})\}) \geq (n+1)^{-L} \exp^{-nH(\pi(\nu^{1,n}) | \pi(\mu^1))}$$

and the analogue for $\mathbb{P}(\{\pi(L_n^Y) = \pi(\nu^{2,n})\})$ also holds true. As a consequence

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in]y_m - \varepsilon/4, y_m + \varepsilon/4]) &\geq \\ &\geq -\limsup(H(\pi(\nu^{1,n})|\pi(\mu^1)) + H(\pi(\nu^{2,n})|\pi(\mu^2))) \\ &\geq -\limsup_{n \rightarrow \infty} (H(\pi(\nu^1)|\pi(\mu^1)) + H(\pi(\nu^2)|\pi(\mu^2))) \end{aligned} \quad (2.6)$$

where (2.5) is due to (2) in Lemma 2.5 and (2.6) comes from the fact that $H(\pi(\nu^1)|\pi(\mu^1)) \leq H(\nu^1|\mu^1)$ for every $\Pi \in \mathcal{P}$, see e.g. Theorem 1.4.3 in [6].

2.3. LD upper bound

We first establish the LD upper bound under the assumption that E is finite. Then we extend this result to the general case.

2.3.1. A particular case

Here we assume that $E = \{a_1, \dots, a_N\}$. For every $n \geq 1$ we have $W_\infty(L_n^X, L_n^Y) \in \Gamma = \{d(a_i, a_j), 1 \leq i, j \leq N\}$ which is a finite set so it is sufficient in order to establish the LD upper bound to show that for every $\delta \in \Gamma$ we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) = \delta) \leq - \inf_{\substack{\nu^1, \nu^2 \in M^1(E) \\ W_\infty(\nu^1, \nu^2) = \delta}} \{H(\nu^1|\mu^1) + H(\nu^2|\mu^2)\}. \quad (2.7)$$

For every $n \geq 1$ we get

$$\begin{aligned} \mathbb{P}(W_\infty(L_n^X, L_n^Y) = \delta) &= \sum_{\substack{\nu^{1,n}, \nu^{2,n} \in M^{1,n}(E) \\ W_\infty(\nu^{1,n}, \nu^{2,n}) = \delta}} \mathbb{P}(L_n^X = \nu^{1,n}, L_n^Y = \nu^{2,n}) \\ &\leq \sum_{\substack{\nu^{1,n}, \nu^{2,n} \in M^{1,n}(E) \\ W_\infty(\nu^{1,n}, \nu^{2,n}) = \delta}} \exp(-nH(\nu^{1,n}|\mu^1) - nH(\nu^{2,n}|\mu^2)) \quad (2.8) \\ &\leq (n+1)^{2N} \exp\left(-n \inf_{\substack{\nu^1, \nu^2 \in M^1(E) \\ W_\infty(\nu^1, \nu^2) = \delta}} \{H(\nu^1|\mu^1) + H(\nu^2|\mu^2)\}\right) \end{aligned}$$

see e.g. Lemma 2.1.9 in [4] for the elementary combinatorial estimate (2.8). The announced (2.7) follows.

2.3.2. The general case

Let us introduce some more notations. For every $\Pi \in \mathcal{P}$ we consider

$$I_\infty^\Pi(x) = \inf_{\substack{\nu^1, \nu^2 \in M^1(E) \\ W_\infty(\nu^1, \nu^2) = x}} \{H(\pi(\nu^1)|\pi(\mu^1)) + H(\pi(\nu^2)|\pi(\mu^2))\}$$

and we shall denote by J_∞^Π the lower semi-continuous regularization of I_∞^Π .

Since $[0, 1]$ is compact the sequence $(W_\infty(L_n^X, L_n^Y))_{n \geq 1}$ is naturally exponentially tight so it is sufficient in order to establish the LD upper bound to consider events of the form $\{W_\infty(L_n^X, L_n^Y) \in [a, b]\}$, see Lemma 1.2.18 in [4]. According to Lemma 2.3 for every $a, b \in [0, 1]$, $a \leq b$, and every $\Pi \in \mathcal{P}$ we have for every $n \geq 1$ that

$$\mathbb{P}(W_\infty(L_n^X, L_n^Y) \in [a, b]) \leq \mathbb{P}(W_\infty(\pi(L_n^X), \pi(L_n^Y)) \in [a - 2\Delta(\Pi), b + 2\Delta(\Pi)]).$$

Hence, due to the computation carried out assuming that E is finite we get for every $\Pi \in \mathcal{P}$

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in [a, b]) &\leq \\ &\leq - \inf_{\substack{\nu^1, \nu^2 \in M^1(\{s_1, \dots, s_L\}) \\ W_\infty(\nu^1, \nu^2) \in [a - 2\Delta(\Pi), b + 2\Delta(\Pi)]}} \{H(\nu^1 | \pi(\mu^1)) + H(\nu^2 | \pi(\mu^2))\} \\ &\leq - \inf_{\substack{\nu^1, \nu^2 \in M^1(E) \\ W_\infty(\nu^1, \nu^2) \in [a - 2\Delta(\Pi), b + 2\Delta(\Pi)]}} \{H(\pi(\nu^1) | \pi(\mu^1)) + H(\pi(\nu^2) | \pi(\mu^2))\} \\ &\leq - \inf_{x \in [a - 2\Delta(\Pi), b + 2\Delta(\Pi)]} I_\infty^\Pi(x) \\ &\leq - \inf_{x \in [a - 2\Delta(\Pi), b + 2\Delta(\Pi)]} J_\infty^\Pi(x). \end{aligned}$$

Hence, to conclude the proof of the LD upper bound we are left to prove that

$$\sup_{\Pi \in \mathcal{P}} \inf_{x \in [a - 2\Delta(\Pi), b + 2\Delta(\Pi)]} J_\infty^\Pi(x) \geq \inf_{x \in [a, b]} J_\infty(x). \quad (2.9)$$

Assume for a while the following

Lemma 2.7. *For every $x \in [0, 1]$ we have $\sup_{\Pi} J_\infty^\Pi(x) = J_\infty(x)$.*

Assume also that (2.9) does not hold : There exists a $\delta > 0$ such that for every $\Pi \in \mathcal{P}$

$$\inf_{x \in [a - 2\Delta(\Pi), b + 2\Delta(\Pi)]} J_\infty^\Pi(x) < \inf_{x \in [a, b]} J_\infty(x) - \delta.$$

Since for every $\Pi \in \mathcal{P}$ the map $x \mapsto J_\infty^\Pi(x)$ is lower semi-continuous there exists a net $(x_\Pi)_{\Pi \in \mathcal{P}}$ such that $x_\Pi \in [a - 2\Delta(\Pi), b + 2\Delta(\Pi)]$ and

$$J_\infty^\Pi(x_\Pi) - \delta/2 < \inf_{x \in [a, b]} J_\infty(x) - \delta.$$

Due to the log sum inequality, see e.g. Theorem 2.7.1 in [3], for every $K, K' \in \mathcal{P}$ such that $K \preceq K'$ and every $x \in [0, 1]$ we have $I_\infty^K(x) \leq I_\infty^{K'}(x)$ hence $J_\infty^K(x) \leq J_\infty^{K'}(x)$ so for every $\Pi' \in \mathcal{P}$ such that $\Pi' \preceq \Pi$ we have

$$J_\infty^{\Pi'}(x_\Pi) - \delta/2 < \inf_{x \in [a, b]} J_\infty(x) - \delta.$$

Since $(x_\Pi)_{\Pi \in \mathcal{P}}$ is a $[0, 1]$ -valued net it admits a converging subnet which limit we denote $x^* \in [a, b]$. Due to the fact that $J_\infty^{\Pi'}$ is lower semi-continuous we get

$$\liminf_{\Pi \in \mathcal{P}} J_\infty^{\Pi'}(x_\Pi) \geq J_\infty^{\Pi'}(x^*)$$

hence

$$J_\infty^{\Pi'}(x^*) < \inf_{x \in [a, b]} J_\infty(x) - \delta/2$$

which, according to Lemma 2.7, implies $J_\infty(x^*) < \inf_{x \in [a, b]} J_\infty(x) - \delta/2$. Since the latter can not be (2.9) holds and the LD upper bound follows.

Proof of Lemma 2.7 As noticed above for every $x \in [0, 1]$ the map $\Pi \mapsto J_\infty^\Pi(x)$ is non-decreasing so $\sup_\Pi J_\infty^\Pi(x) \leq J_\infty(x)$. Conversely let x be a fixed element of $[0, 1]$. We shall assume that $\sup_\Pi J_\infty^\Pi(x) < \infty$ for otherwise the claimed equality trivially holds. Hence for every $\Pi \in \mathcal{P}$ and every $\delta > 0$ there exists $y_{\delta, \Pi}$ such that $|x - y_{\delta, \Pi}| < \delta$, $I_\infty^\Pi(y_{\delta, \Pi}) - \delta/2 < \inf_{y \in B(x, \delta)} I_\infty^\Pi(y)$ and $\lim_{\delta \rightarrow 0} I_\infty^\Pi(y_{\delta, \Pi}) = J_\infty^\Pi(x)$. In particular we can assume that $I_\infty^\Pi(y_{\delta, \Pi}) < \infty$ thus there exists $\nu_1^{\Pi, \delta}, \nu_2^{\Pi, \delta} \in M^1(E)$ such that $W_\infty(\nu_1^{\Pi, \delta}, \nu_2^{\Pi, \delta}) = y_{\delta, \Pi}$ and

$$H(\pi(\nu_1^{\Pi, \delta})|\pi(\mu^1)) + H(\pi(\nu_2^{\Pi, \delta})|\pi(\mu^2)) - \delta \leq I_\infty^\Pi(y_{\delta, \Pi}) - \delta/2. \quad (2.10)$$

Next we define $\rho_1^{\Pi, \delta}, \rho_2^{\Pi, \delta} \in M^1(E)$ by

$$\rho_1^{\Pi, \delta}(F) = \sum_{A_i \in \Pi} \frac{\pi(\nu_1^{\Pi, \delta})(A_i)}{\mu_1(A_i)} \mu_1(F \cap A_i)$$

and accordingly

$$\rho_2^{\Pi, \delta}(F) = \sum_{A_i \in \Pi} \frac{\pi(\nu_2^{\Pi, \delta})(A_i)}{\mu_2(A_i)} \mu_2(F \cap A_i)$$

for every borelian $F \subset E$. Clearly $\pi(\nu_1^{\Pi, \delta}) = \pi(\rho_1^{\Pi, \delta})$ and $\pi(\nu_2^{\Pi, \delta}) = \pi(\rho_2^{\Pi, \delta})$ hence

$$\begin{aligned} |W_\infty(\rho_1^{\Pi, \delta}, \rho_2^{\Pi, \delta}) - W_\infty(\nu_1^{\Pi, \delta}, \nu_2^{\Pi, \delta})| &\leq |W_\infty(\rho_1^{\Pi, \delta}, \rho_2^{\Pi, \delta}) - W_\infty(\pi(\rho_1^{\Pi, \delta}), \pi(\rho_2^{\Pi, \delta}))| + \\ &\quad + |W_\infty(\pi(\rho_1^{\Pi, \delta}), \pi(\rho_2^{\Pi, \delta})) - W_\infty(\pi(\nu_1^{\Pi, \delta}), \pi(\nu_2^{\Pi, \delta}))| + \\ &\quad + |W_\infty(\pi(\nu_1^{\Pi, \delta}), \pi(\nu_2^{\Pi, \delta})) - W_\infty(\nu_1^{\Pi, \delta}, \nu_2^{\Pi, \delta})| \\ &\leq 4\Delta(\Pi) \end{aligned}$$

and

$$\begin{aligned} H(\rho_1^{\Pi, \delta}|\mu^1) + H(\rho_2^{\Pi, \delta}|\mu^2) &= H(\pi(\rho_1^{\Pi, \delta})|\pi(\mu^1)) + H(\pi(\rho_2^{\Pi, \delta})|\pi(\mu^2)) \\ &= H(\pi(\nu_1^{\Pi, \delta})|\pi(\mu^1)) + H(\pi(\nu_2^{\Pi, \delta})|\pi(\mu^2)). \end{aligned}$$

Thus for every $\delta > 0$ and every $\Pi \in \mathcal{P}$ we have.

$$\inf_{u \in B(y_\delta, \Pi, 4\Delta(\Pi))} I_\infty(u) - \delta \leq \inf_{y \in B(x, \delta)} I_\infty^\Pi(y)$$

hence

$$\inf_{y \in B(x, \delta + 4\Delta(\Pi))} I_\infty(y) - \delta \leq \sup_{\Pi \in \mathcal{P}} \sup_{\delta > 0} \inf_{y \in B(x, \delta)} I_\infty^\Pi(y).$$

Since $\Pi \preceq \Pi'$ implies $\Delta(\Pi') \leq \Delta(\Pi)$ we have

$$\sup_{\Pi \in \mathcal{P}} \inf_{y \in B(x, \delta + 4\Delta(\Pi))} I_\infty(y) = \sup_{\substack{\Pi \in \mathcal{P} \\ \Delta(\Pi) < \delta}} \inf_{y \in B(x, \delta + 4\Delta(\Pi))} I_\infty(y)$$

hence

$$\inf_{y \in B(x, 5\delta)} I_\infty(y) \leq \sup_{\Pi \in \mathcal{P}} J_\infty^\Pi(x)$$

whence $\sup_{\Pi} J_\infty^\Pi(x) \geq J_\infty(x)$. \square

3. Additional proofs

Here we give the proofs of the lemmas presented in Section 2.1 and of Proposition 1.1 and 1.2. They are ordered so as to minimize the total length of the paper and not according to the expository order. The proofs of Lemma 2.1 and 2.2 are straightforward generalizations to compact metric spaces of the corresponding results in [2]. We give them for the paper to be self-contained.

3.1. Three more lemmas

We shall employ the following Portmanteau-type result several times in the present section

Lemma 3.1. *Let $(P^j)_{j \in J}$ be a net of Borel probability measures on a metric space (Y, δ) that converges weakly to some probability measure P . For every open $U \subset Y$ we have*

$$\liminf_{j \in J} P^j(U) \geq P(U).$$

Proof of Lemma 3.1. We follow the proof of Theorem 11.1.1 in [5]. Let $F = U^c$. For every $x \in Y$ we define $\delta(x, F) = \inf_{y \in F} \delta(x, y)$ and for every $m \geq 1$, $f_m(x) = \min(1, m\delta(x, F))$ and $F_m = f_m^{-1}(\{1\})$. We see that f_m is an increasing sequence of bounded and continuous functions that converges to 1_U . For every $\varepsilon > 0$ there exists an M_0 such that for every $m \geq M_0$ we have $P(F_m) > P(U) - \varepsilon$ and for those m we have

$$\begin{aligned} \liminf_{j \in J} P^j(U) &\geq \liminf_{j \in J} \int_Y f_m(x) P^j(dx) \\ &= \int_Y f_m(x) P(dx) \\ &\geq P(F_m) \\ &> P(U) - \varepsilon \end{aligned}$$

the conclusion follows by letting $\varepsilon \rightarrow 0$. \square

Lemma 3.2. *Let $(P^j)_{j \in J}$ be a net of Borel probability measures on a compact metric space (Y, δ) that converges weakly to some probability measure P . For every $x \in \mathcal{S}(P)$ and every $j \in J$ there exists an $x^j \in \mathcal{S}(P^j)$ such that the net $(x^j)_{j \in J}$ converges to x .*

Proof of Lemma 3.2. Assume that the announced statement is not true: There exist an $x \in \mathcal{S}(P)$ such that for every net $(x^j)_{j \in J}$ satisfying $x^j \in \mathcal{S}(P^j)$ the net does not converge to x . Consider $(x^j)_{j \in J}$ with x^j defined by $d(x, x^j) = \inf_{y \in \mathcal{S}(P^j)} d(x, y)$. Then there exist an $\varepsilon > 0$ and a cofinal $L \subset J$ such that for every $l \in L$ $d(x, x^l) > \varepsilon$. For every $l \in L$ we have $P^l(B(x, \varepsilon/2)) = 0$ while $P(B(x, \varepsilon/2)) > 0$ by definition. This is impossible according to Lemma 3.1 and the conclusion follows. \square

Lemma 3.3. *Every $P \in M^1(E^2)$ which is the limit in the weak convergence topology of a net $(P^j)_{j \in J}$ of infinitely cyclically monotone elements of $M^1(E^2)$ is infinitely cyclically monotone.*

Proof of Lemma 3.3. Let $n \geq 2$, $(x_1, y_1), \dots, (x_n, y_n) \in \mathcal{S}(P)$ and $\sigma \in \mathfrak{S}_n$. Due to Lemma 3.2 for every $1 \leq i \leq n$ there exists $(x_i^j, y_i^j)_{j \in J}$ such that for every $j \in J$ $(x_i^j, y_i^j) \in \mathcal{S}(P^j)$ and $\lim_{j \in J} (x_i^j, y_i^j) = (x_i, y_i)$. Since for every $j \in J$, P^j is infinitely cyclically monotone we have $\max_{1 \leq i \leq n} d(x_i^j, y_i^j) \leq \max_{1 \leq i \leq n} d(x_i^j, y_{\sigma(i)}^j)$ and passing to the limit we get $\max_{1 \leq i \leq n} d(x_i, y_i) \leq \max_{1 \leq i \leq n} d(x_i, y_{\sigma(i)})$. \square

3.2. Proof of Lemma 2.4

For every $j \in J$ we denote $x^j = \sup \mathcal{S}(P^j)$ and $x = \sup \mathcal{S}(P)$. Let us assume that (2.1) does not hold i.e. there exists an $\eta > 0$ such that $\limsup_{j \in J} x^j < x - \eta$. As a consequence there exists a subnet $(P^l)_{l \in L}$ of $(P^j)_{j \in J}$ such that for every $l \in L$ we have $P^l([x - \eta/2, x + \eta/2]) = 0$ while by definition $P([x - \eta/2, x + \eta/2]) > 0$. But according to Lemma 3.1 we should have $\liminf_{l \in L} P^l([x - \eta/2, x + \eta/2]) \geq P([x - \eta/2, x + \eta/2])$. The conclusion follows. \square

3.3. Proof of Lemma 2.1

Let $\gamma, \nu \in M^1(E)$. We know from Theorem 3.2 in [1] that for every $p \geq 1$ there exists $P^p \in (\gamma, \nu)$ such that

$$\left(\int d(x, y)^p P^p(dx, dy) \right)^{1/p} = \inf_{Q \in \mathcal{C}(\gamma, \nu)} C_p(Q)$$

where

$$C_p(Q) = \left(\int d(x, y)^p Q(dx, dy) \right)^{1/p}.$$

Moreover, P^p is p -cyclically monotone: For every $n \geq 2$, every $(x_1, y_1), \dots, (x_n, y_n) \in \mathcal{S}(P^p)$ and every $\sigma \in \mathfrak{S}_n$ we have

$$\left(\sum_{i=1}^n d(x_i, y_i)^p \right)^{1/p} \leq \left(\sum_{i=1}^n d(x_i, y_{\sigma(i)})^p \right)^{1/p}.$$

Since $M^1(E^2)$ is compact for the weak convergence topology the sequence $(P^p)_{p \geq 1}$ admits a converging sub-sequence which limit we denote $P \in \mathcal{C}(\gamma, \nu)$. We shall prove that

$$\sup \mathcal{S}(P \circ d^{-1}) = \inf_{Q \in \mathcal{C}(\gamma, \nu)} \sup \mathcal{S}(Q \circ d^{-1}) \quad (3.1)$$

and that P is infinitely cyclically monotone. Indeed, for any $Q \in (\gamma, \nu)$ and any $1 \leq q \leq p$ due to the optimality of P^p and Holder's inequality we have that

$$C_q(P^p) \leq C_p(P^p) \leq C_p(Q).$$

Taking $p \rightarrow \infty$ then $q \rightarrow \infty$ we obtain $C_\infty(P) \leq C_\infty(Q)$ where

$$C_\infty(Q) = \sup \mathcal{S}(Q \circ d^{-1}),$$

which proves (3.1). Now let $n \geq 2$, $(x_1, y_1), \dots, (x_n, y_n) \in \mathcal{S}(P)$ and $\sigma \in \mathfrak{S}_n$. According to Lemma 3.2 for every $1 \leq i \leq n$ and every $p \geq 1$ there exists $(x_i^p, y_i^p) \in \mathcal{S}(P^p)$ such that $(x_i^p, y_i^p) \rightarrow (x_i, y_i)$ as $p \rightarrow \infty$. Since every P^p is p -cyclically monotone we have

$$\left(\sum_{i=1}^n d(x_i, y_i)^p \right)^{1/p} \leq \left(\sum_{i=1}^n d(x_i, y_{\sigma(i)})^p \right)^{1/p}.$$

Taking the limit $p \rightarrow \infty$ yields $\max_{1 \leq i \leq n} d(x_i, y_i) \leq \max_{1 \leq i \leq n} d(x_i, y_{\sigma(i)})$ which concludes the proof. \square

3.4. Proof of Lemma 2.2

Let P be an infinitely cyclically monotone element of $M^1(E^2)$. let us assume that there exist a $Q \in \mathcal{C}(P_1, P_2)$ and $\varepsilon > 0$ such that

$$\sup \mathcal{S}(P \circ d^{-1}) \geq 10\varepsilon + \sup \mathcal{S}(Q \circ d^{-1}). \quad (3.2)$$

Since E is compact there exists a finite family $(c_i)_{1 \leq i \leq k}$ of elements of E such that $E \subset \cup_{i=1}^k B(c_i, \varepsilon)$. We introduce $C = \{c_1, \dots, c_k\}$, $V_1 = B(c_1, \varepsilon)$ and for every $2 \leq j \leq k$, $V_j = B(c_j, \varepsilon) \cap (\cup_{i=1}^{j-1} V_i)^c$. We further define $\gamma, \tilde{\gamma} \in M^1(E^2)$ by

$$\gamma = \sum_{i,j=1}^k P(V_i \times V_j) \delta_{(c_i, c_j)} \text{ and } \tilde{\gamma} = \sum_{i,j=1}^k Q(V_i \times V_j) \delta_{(c_i, c_j)}.$$

Since P and Q have the same marginals so do γ and $\tilde{\gamma}$ and one has

$$(x, y) \in \mathcal{S}(\gamma) \Rightarrow \text{there exists } \tilde{x} \in C \text{ such that } (\tilde{x}, y) \in \mathcal{S}(\tilde{\gamma}) \quad (3.3)$$

and

$$(\tilde{x}, \tilde{y}) \in \mathcal{S}(\tilde{\gamma}) \Rightarrow \text{there exists } y \in C \text{ such that } (\tilde{x}, y) \in \mathcal{S}(\gamma). \quad (3.4)$$

Due to (3.2) there exists some $(x_0, y_0) \in \mathcal{S}(\gamma)$ such that

$$d(x_0, y_0) \geq 5\varepsilon + \max \mathcal{S}(\tilde{\gamma} \circ d^{-1}) \quad (3.5)$$

and since P is infinitely cyclically monotone we also have for every $(x_1, y_1), \dots, (x_n, y_n) \in \mathcal{S}(\gamma)$ and every $\sigma \in \mathfrak{S}_n$ that

$$\max_{1 \leq i \leq n} d(x_i, y_i) \leq 4\varepsilon + \max_{1 \leq i \leq n} d(x_i, y_{\sigma(i)}). \quad (3.6)$$

Due to (3.3) and (3.4) we recursively define two sequences $(\mathcal{D}_m)_{m \geq 1}$ and $(\mathcal{E}_m)_{m \geq 0}$ of subsets of C by setting $\mathcal{E}_0 = \{y_0\}$ and for $m \geq 1$,

$$\mathcal{D}_m = \{\tilde{x} : \text{there exists } y \in \mathcal{E}_{m-1} \text{ such that } (\tilde{x}, y) \in \mathcal{S}(\tilde{\gamma})\}$$

and

$$\mathcal{E}_m = \{y : \text{there exists } \tilde{x} \in \mathcal{D}_m \text{ such that } (\tilde{x}, y) \in \mathcal{S}(\gamma)\}.$$

We further define $\mathcal{D} = \cup_{m \geq 1} \mathcal{D}_m$ and $\mathcal{E} = \cup_{m \geq 0} \mathcal{E}_m$. There are two alternatives: either x_0 belongs to \mathcal{D} or not. They both lead to a contradiction.

First case: $x_0 \in \mathcal{D}$. In this case there exists $m \geq 1$ such that $x_0 \in \mathcal{D}_m$ and by going backwards from \mathcal{D}_m to \mathcal{E}_0 it is possible to define two families $(x_i)_{0 \leq i \leq m}$ and $(y_i)_{0 \leq i \leq m-1}$ such that for all $i \in \{0, \dots, m-1\}$

$$(x_i, y_i) \in \mathcal{S}(\gamma) \text{ and } (x_{i+1}, y_i) \in \mathcal{S}(\tilde{\gamma}),$$

where we set $x_m = x_0$. Due to (3.6) we have

$$\max_{0 \leq i \leq m-1} d(x_i, y_i) - 4\varepsilon \leq \max_{0 \leq i \leq m-1} d(x_{i+1}, y_i).$$

Since $\max_{0 \leq i \leq m-1} d(x_i, y_i) \geq d(x_0, y_0)$, due to (3.5) we get

$$\max_{(x,y) \in \mathcal{S}(\tilde{\gamma})} d(x, y) + \varepsilon \leq \max_{0 \leq i \leq m-1} d(x_{i+1}, y_i)$$

but for every $i \in \{0, \dots, m-1\}$ we have $(x_{i+1}, y_i) \in \mathcal{S}(\tilde{\gamma})$ which yields a contradiction.

Second case: $x_0 \in \mathcal{D}^c$. From the definitions of \mathcal{D} and \mathcal{E} we notice the following two facts

$$x \in \mathcal{D} \text{ and } (x, y) \in \mathcal{S}(\gamma) \Rightarrow y \in \mathcal{E} \quad (3.7)$$

and

$$\tilde{y} \in \mathcal{E} \text{ and } (\tilde{x}, \tilde{y}) \in \mathcal{S}(\tilde{\gamma}) \Rightarrow \tilde{x} \in \mathcal{D}. \quad (3.8)$$

Due to this and since γ and $\tilde{\gamma}$ have the same marginals one has

$$\gamma(\mathcal{D} \times \mathcal{E}) = \gamma(\mathcal{D} \times C) = \tilde{\gamma}(\mathcal{D} \times C)$$

and

$$\tilde{\gamma}(\mathcal{D} \times \mathcal{E}) = \tilde{\gamma}(C \times \mathcal{E}) = \gamma(C \times \mathcal{E}),$$

hence

$$\gamma(\mathcal{D} \times \mathcal{E}) = \tilde{\gamma}(\mathcal{D} \times C) \geq \tilde{\gamma}(\mathcal{D} \times \mathcal{E}) = \gamma(C \times \mathcal{E}).$$

This implies that $\gamma((C \cap \mathcal{D}^c) \times \mathcal{E}) = 0$ while $(x_0, y_0) \in (C \cap \mathcal{D}^c) \times \mathcal{E}$ and $\gamma(\{(x_0, y_0)\}) > 0$ which constitute a contradiction. \square

3.5. Proof of Lemma 2.3

Let $\nu^1, \nu^2 \in M^1(E)$ and $\Pi = (A_1, \dots, A_L) \in \mathcal{P}$. According to Lemma 2.1 there exists $P \in \mathcal{C}(\nu^1, \nu^2)$ such that $W_\infty(\nu^1, \nu^2) = \sup \mathcal{S}(P \circ d^{-1})$. Consider $P^\Pi \in M^1(\{s_1, \dots, s_L\}^2)$ defined by

$$P^\Pi = \sum_{i,j=1}^L P(A_i \times A_j) \delta_{(s_i, s_j)}.$$

Since $P \in \mathcal{C}(\nu^1, \nu^2)$ necessarily $P^\Pi \in \mathcal{C}(\pi(\nu^1), \pi(\nu^2))$ hence

$$\begin{aligned} W_\infty(\pi(\nu^1), \pi(\nu^2)) &\leq \sup \mathcal{S}(P^\Pi \circ d^{-1}) \\ &\leq \sup \mathcal{S}(P \circ d^{-1}) + 2\Delta(\Pi) \\ &\leq W_\infty(\nu^1, \nu^2) + 2\Delta(\Pi). \end{aligned}$$

We are left to prove that $W_\infty(\nu^1, \nu^2) \leq W_\infty(\pi(\nu^1), \pi(\nu^2)) + 2\Delta(\Pi)$. Again, according to Lemma 2.1 there exists $Q^\Pi \in \mathcal{C}(\pi(\nu^1), \pi(\nu^2))$ such that $W_\infty(\pi(\nu^1), \pi(\nu^2)) = \sup \mathcal{S}(Q^\Pi \circ d^{-1})$. Clearly Q^Π is of the form

$$Q^\Pi = \sum_{i,j=1}^L Q_{i,j} \delta_{(s_i, s_j)}.$$

Now we consider $Q \in M^1(E^2)$ defined by

$$Q(F) = \sum_{i,j=1}^L \frac{Q_{i,j}}{\nu^1(A_i) \nu^2(A_j)} \nu^1 \otimes \nu^2(F \cap (A_i \times A_j)). \quad (3.9)$$

Since $Q^\Pi \in \mathcal{C}(\pi(\nu^1), \pi(\nu^2))$ we see that $Q_{i,j} = 0$ as soon as $\nu^1(A_i) = 0$ or $\nu^2(A_j) = 0$ so Q is well defined and we can assume that the sum in (3.9) runs

over i, j such that $\nu^1(A_i)\nu^2(A_j) \neq 0$. Let us check that e.g. $Q_1 = \nu^1$. For every measurable $U \subset E$ we have

$$\begin{aligned}
Q(U \times E) &= \sum_{i,j=1}^L \frac{Q_{i,j}}{\nu^1(A_i)\nu^2(A_j)} \nu^1 \otimes \nu^2((U \times E) \cap (A_i \times A_j)) \\
&= \sum_{i,j=1}^L \frac{Q_{i,j}}{\nu^1(A_i)\nu^2(A_j)} \nu^1(U \cap A_i) \nu^2(A_j) \\
&= \sum_{i,j=1}^L \frac{Q_{i,j}}{\nu^1(A_i)} \nu^1(U \cap A_i) \\
&= \sum_{i=1}^L \frac{1}{\nu^1(A_i)} \left(\sum_{j=1}^L Q_{i,j} \right) \nu^1(U \cap A_i) \\
&= \sum_{i=1}^L \nu^1(U \cap A_i) \\
&= \nu^1(U)
\end{aligned} \tag{3.10}$$

where (3.10) is due to the fact that $Q^\Pi \in \mathcal{C}(\pi(\nu^1), \pi(\nu^2))$. Now

$$\begin{aligned}
W_\infty(\nu^1, \nu^2) &\leq \sup \mathcal{S}(Q \circ d^{-1}) \\
&\leq \sup \mathcal{S}(Q^\Pi \circ d^{-1}) + 2\Delta(\Pi) \\
&\leq W_\infty(\pi(\nu^1), \pi(\nu^2)) + 2\Delta(\Pi).
\end{aligned}$$

which concludes the proof. \square

3.6. Proof of Proposition 1.1

First let us prove that if $x \in \mathcal{Z}_\mu$ then $I_\infty(x) = 0$. We can assume that $x \neq 0$ since we clearly have $I_\infty(0) = 0$. So let $x \in \mathcal{Z}_\mu$ be such that $x \neq 0$: There exist $a_i, a_j \in \mathcal{S}(\mu)$ such that $a_i \neq a_j, a_i \leftrightarrow a_j$ and $d(a_i, a_j) = x$. Since $\mathcal{S}(\mu)$ is finite there exists an $\varepsilon_0 > 0$ such that for every $0 < \varepsilon < \varepsilon_0$ the probability measure

$$\nu^\varepsilon(a_l) = \begin{cases} \mu(a_l) & \text{if } l \neq i \text{ and } l \neq j \\ \mu(a_i) - \varepsilon & \text{if } l = i \\ \mu(a_j) + \varepsilon & \text{if } l = j \end{cases}$$

is well-defined. Now we prove that for every $0 < \varepsilon < \varepsilon_0$ we have $W_\infty(\nu^\varepsilon, \mu) = d(a_i, a_j) = x$. Indeed,

$$Q^\varepsilon = \sum_{\substack{l=1 \\ l \neq i}}^n \mu(a_l) \delta_{(a_l, a_l)} + (\mu(a_i) - \varepsilon) \delta_{(a_i, a_i)} + \varepsilon \delta_{(a_i, a_j)}$$

satisfies $Q_1^\varepsilon = \mu$ and $Q_2^\varepsilon = \nu^\varepsilon$ and it is infinitely cyclically monotone as we prove now. Let $n \geq 2$, $(x_1, y_1), \dots, (x_n, y_n)$ and $\sigma \in \mathfrak{S}_n$. If all the (x_i, y_i) 's are from the diagonal then obviously $\max_{1 \leq i \leq n} d(x_i, y_i) \leq \max_{1 \leq i \leq n} d(x_i, y_{\sigma(i)})$. So let us assume that e.g. $(x_1, y_1) = (a_i, a_j)$ with all the other (x_i, y_i) 's from the diagonal. In this case $\max_{1 \leq i \leq n} d(x_i, y_i) = d(a_i, a_j)$. The permutation σ defines a transport from a_i to a_j in the following way: We start at $x_1 = a_i$ and move to $y_{\sigma(1)}$. The latter is either a_j in which case $d(a_i, a_j) \leq \max_{1 \leq i \leq n} d(x_i, y_{\sigma(i)})$ and the proof that Q^ε is infinitely cyclically monotone is completed or it corresponds to the second coordinate of a diagonal term (x_l, y_l) . Next from $y_{\sigma(1)} = y_l = x_l$ we move to $y_{\sigma(l)}$. Again, we stop if $y_{\sigma(l)} = a_j$ or continue like that until we finally arrive at a_j . Since $a_i \leftrightarrow a_j$ we know that along this travel there is a stage from some x_s to $y_{\sigma(s)}$ such that $d(a_i, a_j) \leq d(x_s, y_{\sigma(s)}) \leq \max_{1 \leq i \leq n} d(x_i, y_{\sigma(i)})$ so Q^ε is infinitely cyclically monotone. Finally, for every $0 < \varepsilon < \varepsilon_0$ we have that $I_\infty(x) \leq H(\nu^\varepsilon | \mu) + H(\mu | \mu)$ and we obtain $I_\infty(x) = 0$ by letting $\varepsilon \rightarrow 0$.

Now we prove that if $x \in Z_\mu^c$ then $I_\infty(x) > 0$. If $x \notin \Gamma = \{d(a_i, a_j), 1 \leq i, j \leq N\}$ we clearly have $I_\infty(x) = \infty$. So let us consider $x \in \Gamma \cap Z_\mu^c$ and assume that $I_\infty(x) = 0$: There exists a sequence $(\nu^{1,n}, \nu^{2,n})_{n \geq 1}$ of elements of $M^1(E)$ such that for every $n \geq 1$ we have $W_\infty(\nu^{1,n}, \nu^{2,n}) = x$ and

$$H(\nu^{1,n} | \mu) + H(\nu^{2,n} | \mu) < \frac{1}{n}. \quad (3.11)$$

Since $M^1(E)$ is compact we can extract from $(\nu^{1,n}, \nu^{2,n})_{n \geq 1}$ a sub-sequence that we still denote $(\nu^{1,n}, \nu^{2,n})_{n \geq 1}$. Due to 3.11 we necessarily have $\nu^{1,n} \xrightarrow{w} \mu$ and $\nu^{2,n} \xrightarrow{w} \mu$. According to Lemma 2.1 for every $n \geq 1$ there exists an infinitely cyclically monotone $P^n \in \mathcal{C}(\nu^{1,n}, \nu^{2,n})$ such that $\sup \mathcal{S}(P^n \circ d^{-1}) = x$. Since $x \in \Gamma \cap Z_\mu^c$ there exists an $(a_t, a_s) \in \mathcal{S}(\mu)^2$ such that $x = d(a_t, a_s)$, $a_t \not\leftrightarrow a_s$ and we assume that for every $n \geq 1$ $(a_t, a_t) \in \mathcal{S}(P^n)$ (we can consider a sub-sequence of P^n if needed). Again, since $M^1(E^2)$ is compact we can extract from $(P^n)_{n \geq 1}$ a converging sub-sequence which limit we denote P . Necessarily $P \in \mathcal{C}(\mu, \mu)$ and according to Lemma 3.3 P must be infinitely cyclically monotone hence $P(a_i, a_i) = \mu(a_i)$ for every $1 \leq i \leq N$ while $P(a_i, a_j) = 0$ whenever $i \neq j$. So, there exist a $\delta > 0$ and an N_0 such that for every $n \geq N_0$ we have $\min_{1 \leq l \leq N} \{P^n(a_i, a_i)\} > \delta$. The latter implies that $(x_1, y_1) = (a_1, a_1), \dots, (x_N, y_N) = (a_N, a_N), (x_{N+1}, y_{N+1}) = (a_t, a_s) \in \mathcal{S}(P^n)$ so for every $\sigma \in \mathfrak{S}_{N+1}$

$$d(a_t, a_s) = \max_{1 \leq i \leq N+1} d(x_i, y_i) \leq \max_{1 \leq i \leq N+1} d(x_i, y_{\sigma(i)})$$

since P^n is infinitely cyclically monotone but this contradicts $a_t \not\leftrightarrow a_s$ and the conclusion follows. \square

3.7. Proof of Proposition 1.2

Clearly $J_\infty(0) = 0$. Next let us assume that there exists an $x \in]0, 1]$ such that $J_\infty(x) = 0$. Due to the LD lower bound, for every ε such that $0 < 4\varepsilon < x$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in [x - \varepsilon, x + \varepsilon]) \geq 0.$$

hence

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in [x - \varepsilon, x + \varepsilon]) = 0. \quad (3.12)$$

Next consider $\Pi = (A_1, \dots, A_L) \in \mathcal{P}$ the partitioning of E induced by the grid partitioning of \mathbb{R}^d along its axis into cells of diameter $\varepsilon/2$. In particular we can assume that the associated $\{s_1, \dots, s_L\}$ are such that for any two neighboring cells A_i, A_j we have $d(s_i, s_j) \leq \varepsilon$. According to Lemma 2.3 we have

$$\{W_\infty(L_n^X, L_n^Y) \in [x - \varepsilon, x + \varepsilon]\} \subset \{W_\infty(\pi(L_n^X), \pi(L_n^Y)) \in [x - 3\varepsilon, x + 3\varepsilon]\}.$$

Since $x - 3\varepsilon > \varepsilon$, according to Proposition 1.1 we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(W_\infty(\pi(L_n^X), \pi(L_n^Y)) \in [x - 3\varepsilon, x + 3\varepsilon]) < 0$$

which contradicts (3.12) and concludes the proof. \square

Acknowledgements

I thank Guillaume Carlier, Luigi De Pascale and Christian Léonard for valuable discussions on this problem.

References

- [1] AMBROSIO, L. and PRATELLI, A. (2003). Existence and stability results in the L^1 theory of optimal transportation. *Lecture Notes in Math.* **1813** 123–160. MR2006307
- [2] CHAMPION, TH., DE PASCALE, L. and JUUTINEN, P. (2008). The ∞ -Wasserstein distance: local solutions and existence of optimal transport maps. *SIAM J. Math. Anal.* **40** 1–20. MR2403310
- [3] COVER, TH. and THOMAS, J. A. (1992). *Elements of information theory*. John Wiley & Sons Inc., New York. MR1122806
- [4] DEMBO, A. and ZEITOUNI, O. (1998). *Large deviations techniques and applications*, 2nd ed. Springer-Verlag, New York. MR1619036
- [5] DUDLEY, R. M. (2002). *Real analysis and probability*. Cambridge University Press, Cambridge. MR1932358
- [6] DUPUIS, P. and ELLIS, R. S. (1997). *A weak convergence approach to the theory of large deviations*. A Wiley-Interscience Publication, New York. MR1431744

- [7] GANESH, A. and O'CONNELL, N. (2007). Large and moderate deviations for matching problems and empirical discrepancies. *Markov Process. Related Fields* **13** 85–98. MR2321752
- [8] LEIGHTON, T. and SHOR, P. (1989). Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. *Combinatorica* **9** 161–187. MR1030371
- [9] MOORE, T. O. (1964). *Elementary general topology*. Prentice-Hall Mathematics Series, Englewood Cliffs, N.J.. MR0166745
- [10] ROCKAFELLAR, R. T. and WETS, R. (1998). *Variational analysis*. Grundlehren der Mathematischen Wissenschaften **317**. Springer-Verlag, Berlin. MR1491362
- [11] SERRE, D. (2002). *Matrices*. Graduate Texts in Mathematics **216**. Springer-Verlag, New-York. MR1923507
- [12] TALAGRAND, M. (1994). Matching theorems and empirical discrepancy computations using majorizing measures *J. Amer. Math. Soc.* **7** 455–537. MR1227476
- [13] VILLANI, C. (2009). *Optimal transport, Old and New*. Grundlehren der Mathematischen Wissenschaften **338**. Springer-Verlag, Berlin. MR2459454
- [14] WANG, R., WANG, X. and WU, L. (2010). Sanov's theorem in the Wasserstein distance: a necessary and sufficient condition. *Statist. Probab. Lett.* **80** 505–512. MR2593592